

DEPARTMENT OF INFORMATION TECHNOLOGY

QUESTION BANK

SUB NAME: INFORMATION MANAGEMENT

SUB CODE: IT6701

UNIT I

PART A

1. Define data modeling.

Data modeling is often the first step in database design and object-oriented programming as the designers first create a conceptual model of how data items relate to each other. Data modeling involves a progression from conceptual model to logical model to physical schema.

2. List the types of data models.

- Object Based Data Models
- Physical Data Models
- Record Based Data Models

3. Define schemas.

The term "schema" refers to the organization of data as a blueprint of how the database is constructed (divided into database tables in the case of relational databases). The formal definition of a database schema is a set of formulas (sentences) called integrity constraints imposed on a database.

4. Where physical model is used?

Physical model (most commonly referred to simply as a model but in this context distinguished from a [conceptual model](#)) is a smaller or larger physical copy of an object. The object being modelled may be small (for example, an atom) or large (for example, the [Solar System](#)).

5. Name any two sources of Business rules.

A business rule is a rule that defines or constrains some aspect of [business](#) and always resolves to either true or false. Business rules are intended to assert business structure or to control or influence the behavior of the business.^[1] Business rules describe the operations, definitions and constraints that apply to an organization. Business rules can apply to people, processes, corporate behavior and computing systems in an organization, and are put in place to help the organization achieve its goals.

6. Define Business Rules with an example.

License inspection project

Rule 1-A driver of a vehicle must have a valid driver's license

Rule 2-A driver's license must be considered valid:-

- The driver's license belongs to the driver
- The expiry date of the driver's license is later than the inspection date

7. Summarize the functionalities of JDBC.

- [Statement](#)
- [PreparedStatement](#)
- [CallableStatement](#)

8. Discuss OLEDB.

OLE DB (Object Linking and Embedding Database) is Microsoft's strategic low-level application program interface ([API](#)) for access to different data sources. OLE DB includes not only the Structured Query Language ([SQL](#)) capabilities of the Microsoft-sponsored standard data interface Open Database Connectivity ([ODBC](#)) but also includes access to [data](#) other than SQL data.

9. Illustrate stored procedure with an example.

A stored procedure is a set of Structured Query Language (SQL) statements with an assigned name that's stored in the database in compiled form so that it can be shared by a number of programs.

10. Demonstrate ACID properties.

When a transaction processing system creates a transaction, it will ensure that the transaction will have certain characteristics. The developers of the components that comprise the transaction are assured that these characteristics are in place. They do not need to manage these characteristics themselves. These characteristics are known as the ACID properties. ACID is an acronym for atomicity, consistency, isolation, and durability.

11. Discover Map parameters.

- A mapping parameter represents a constant value that we can define before running a session.
- A mapping parameter retains the same value throughout the entire session.

12. Analyse Hadoop as a Service.

Hadoop as a Service, as offered by Qubole Data Service (QDS) is a cloud computing solution that makes medium and large-scale data processing accessible, easy, fast and inexpensive. This is achieved by eliminating the operational challenges of running Hadoop, so you can focus on business growth.

13. Point out any 2 features of Hadoop Cluster.

1. Scalable-Hadoop is a highly scalable storage platform, because it can store and distribute very large data sets across hundreds of inexpensive servers that operate in parallel. Unlike traditional relational database systems (RDBMS) that can't scale to process large amounts of data, Hadoop enables businesses to run applications on thousands of nodes involving thousands of terabytes of data.

2. Fast-Hadoop's unique storage method is based on a distributed file system that basically 'maps' data wherever it is located on a cluster. The tools for data processing are often on the same servers where the data is located, resulting in much faster data processing. If you're dealing with large volumes of unstructured data, Hadoop is able to efficiently process terabytes of data in just minutes, and petabytes in hours.

14. Compare Map Stage and Reduce stage.

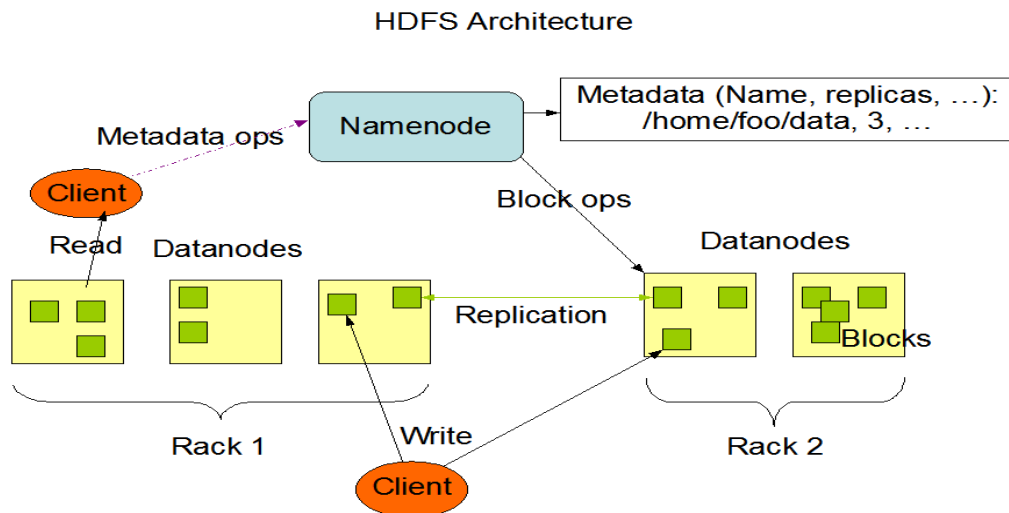
MapReduce is a programming model and an associated implementation for processing and generating large data sets with a parallel, distributed algorithm on a cluster. Conceptually similar approaches have been very well known since 1995 with the Message Passing Interface standard having reduce and scatter operations.

15. Explain the features of Hive.

1. Fits the low level interface requirement of Hadoop perfectly.
2. Supports external tables which make it possible to process data without actually storing in HDFS.
3. It has a rule based optimizer for optimizing logical plans.
4. Supports partitioning of data at the level of tables to improve performance.

16. Design HDFS architecture.

NameNode and DataNodes. HDFS has a master/slave architecture. An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients



PART B

1. Tell about the JDBC in detail.
2. Explain the following SELECT statement syntax with examples in HiveQL.
 - i) Computing with Columns (4)
 - ii) WHERE Clauses (4)
 - iii) GROUP BY Clauses(4)
 - iv) HAVING Clauses (4)
3. List the various data models in database design
4. i) Define NoSQL and is it the next big trend in databases?(8)
ii) Tabulate SQL vs NoSQL(8)
5. Discuss MapR converged data platform(16)
6. Analyse various databases used in NoSQL.(16)
7. Explain Hive unit testing frameworks
8. Develop a program to establish Java Database connectivity

UNIT II

PART A

1. Define malicious code.

Malicious code is the term used to describe any code in any part of a software system or script that is intended to cause undesired effects, security breaches or damage to a system. Malicious code is an application security threat that cannot be efficiently controlled by conventional antivirus software alone.

2. Define TOCTTOU errors.

Time of check to time of use is a class of [software bug](#) caused by changes in a system between the checking of a condition (such as a security credential) and the use of the results of that check. This is one example of a [race condition](#).

3. Define firewall.

A firewall is a [network](#) security system, either hardware- or software-based, that controls incoming and outgoing network traffic based on a set of rules.

4. Define Trapdoors.

Trap doors, also referred to as **backdoors**, are bits of code embedded in programs by the programmer(s) to quickly gain access at a later time, often during the testing or debugging phase. If an unscrupulous programmer purposely leaves this code in or simply forgets to remove it, a potential security hole is introduced. Hackers often plant a backdoor on previously compromised systems to gain later access. Trap doors can be almost impossible to remove in a reliable manner. Often, reformatting the system is the only sure way.

5. Define Salami attack.

A salami attack is when small attacks add up to one major attack that can go undetected due to the nature of this type of cyber crime. It also known as salami slicing.

6. Give the types of Security flaws.

- Memory safety violations,
- Input validation errors
- Race conditions
- Privilege-confusion bugs
- Privilege escalation.

7. Describe the kinds of buffer overflows

There are two main types of buffer overflow attacks: stack based and heap based. Heap-based attacks flood the memory space reserved for a program, but the difficulty

involved with performing such an attack makes them rare. Stack-based buffer overflows are by far the most common.

8. Give the various forms of Malicious code.

- Viruses: pieces of code that attach to host programs and propagate when an infected program executes
- Worms: particular to networked computers, carry out pre-programmed attacks to jump across the network
- Trojan Horses: hide malicious intent inside a host program that appears to do something useful
- Attack scripts: programs written by experts to exploit security weaknesses, usually across the network
- Java attack applets: programs embedded in Web pages that gain foothold through a browser
- ActiveX controls: program components that allow malicious code fragment to control applications or the OS

9. Illustrate Signature based detection.

Signature-based detection really is more along the lines of intrusion detection than firewalls. However, many personal firewalls and some corporate firewalls contain this functionality. Essentially, the system can be configured to look for specific patterns, known to be malicious, and block the traffic. This would help with something like the Code Red worm.

10. Demonstrate intrusion detection system.

An **intrusion detection system** (IDS) is a device or software application that monitors network or **system** activities for malicious activities or policy violations and produces electronic reports to a management station.

11. Point out 3 types of filtering mechanisms in firewall.

- Packet filters
- proxy filters
- stateful packet filters

12. Explain Vulnerability assessment.

Vulnerability analysis, also known as vulnerability assessment, is a process that defines, identifies, and classifies the security holes (vulnerabilities) in a computer, network, or communications infrastructure.

PART B

1. Tell in detail about
 - i) Incomplete Mediation (10)
 - ii) TOCTTOU errors (6)

2. Explain the following
 - i) Software firewalls (8)
 - ii) Hardware firewalls (8)

3.
 - i) List the kinds of Malicious code (8)
 - ii) Prevention of virus infection (8)

4. Briefly describe the following in Developmental Controls
 - i. Peer Reviews (6)
 - ii. Hazard Analysis (4)
 - iii. Testing (6)

5.
 - i) Summarize the various firewall rules (8)
 - ii) Explain HIDS and NIDS in detail (8)

6. Describe the following
 - i) Buffer overflows Flaws (8)
 - ii) Faults (4)
 - iii) Failures (4)

7. Discuss the following in controls against threats(16)
 - i) Modularity (6)
 - ii) Encapsulation (4)
 - iii) Information hiding (6)

8. Illustrate the data protection principles in detail(16)

UNIT III

PART – A

1. Define Master Data

Master data is the core data that is essential to operations in a specific business or business unit. The kinds of information treated as master data varies from one industry to another and even from one company to another within the same industry.

2. List some of the Master Data Domains.

- Manufacturing domains
- Healthcare domains
- Financial services domains
- Education domains

3. Define Collaborative Authoring.

collaborative authoring refers to projects where written works are created by multiple people together (collaboratively) rather than individually. Some projects are overseen by an editor or editorial team, but many grow without any oversight.

4. Discuss various MDM implementation styles.

- Registry Style
- Consolidation Style
- Coexistence Style
- Transaction Style
- Adaptive Transaction Style
- Confederation Style

5. Name any two MDM Architecture patterns.

An enterprise and business architecture would need to integrate information and re-use across business functions and across the technology architecture. Party management capabilities - any legal entity in the world

- A customer master
- A supply chain master

A single reference data source would be ideal for an effective EA and BA pattern. The offer development patterns if we use APQC business process framework as a reference data source we could benefit from the known dependencies between offer development and market/sell offers.

6. Define Data Governance.

Data governance (DG) refers to the overall management of the availability, usability, integrity, and security of the data employed in an enterprise. A sound data governance program includes a governing body or council, a defined set of procedures, and a plan to execute those procedures.

7. Describe universal goals of Data Governance.

1. Enable better decision-making
2. Reduce operational friction
3. Protect the needs of data stakeholders
4. Train management and staff to adopt common approaches to data issues
5. Build standard, repeatable processes
6. Reduce costs and increase effectiveness through coordination of efforts
7. Ensure transparency of processes

8. Discuss Data Governance focus areas.

- They all have activities that address a three-part governance mission: to create rules, resolve conflicts, and provide ongoing services.
- They all employ most or all of the universal components of a Data Governance program.
- They all address universal governance processes and services, such as Issue Resolution and Stakeholder Care.

9. Illustrate Data Governance Organization roles.

First, a group of individuals (or a hierarchy of groups) representing a cross-section of stakeholder groups makes a set of rules in the form of policies, standards, requirements, guidelines, or data definitions. (Or, they gather and align rules. Or address gaps and overlaps in rule sets. Or interpret rules. Or establish guidelines for how to layer rules on top of each other.)

Next, Data Governance always includes a mechanism for resolving data-related issues. Issues are generally addressed at several levels, with a clear escalation path. A particular issue, then, may be resolved by an individual Data Steward, a Stewardship working group, the entire Data Stewardship Council, or the highest-level Data Governance Board.

10. Illustrate some common data governance processes.

1. Aligning Policies, Requirements, and Controls
2. Establishing Decision Rights
3. Establishing Accountability
4. Performing Stewardship
5. Managing Change
6. Defining Data
7. Resolving Issues
8. Specifying Data Quality Requirements
9. Building Governance Into Technology
10. Stakeholder Care
11. Communications
12. Measuring and Reporting Value

11. Examine the maturity levels for Data Risk Management Framework.

- identification of weaknesses
- identification of strengths
- definition of a prioritized measures list
- assessment of the maturity level of ERM
- documentation of the ERM
- overview of results considering different dimensions
- multi-period assessment on different reference dates
- overview of multi-period results considering different dimensions.

12. Analyse Governance Maturity levels.

- Sub-standard
- In development
- essential
- pro active
- transformational

13. Point out the need of Information Security.

- how you select appropriate techniques to tackle and solve problems in the discipline of information security management;
- why security and its management are important for any modern organisation;
- How an information security management system should be planned, documented, implemented and improved, according to the BSi standard on information security management

14. Assess the Google Privacy policy.

- What information we collect and why we collect it.
- How we use that information.
- The choices we offer, including how to access and update information.

PART – B

1. Tell the following about Manage Data in detail
 - i) Process Description (8)
 - ii) Control Objectives (8)

2. Discuss the following in Maturity Model of Data Governance
 - i) Undisciplined (8)
 - ii) Reactive (8)

3. Describe the following
 - i) Manage Data – Management Guidelines (8)
 - ii) Manage Data – Maturity Model (8)

4. Identify the need for Master Data Management and give the solutions for it(16)

5. Summarize the following in detail
 - i) Data Governance Framework (8)
 - ii) Data Governance Lifecycle (8)

6. Describe the following MDM Implementation Styles in detail
 - i) Consolidation (4)
 - ii) Registry (4)
 - iii) Coexistence (4)
 - iv) Transaction (4)

7. Discuss the following in detail
 - i) MDM Conceptual Architecture (8)
 - ii) MDM Logical Architecture (8)

8. Demonstrate the conceptions of privacy in detail.

9. Explain the following stages in Data Governance Maturity Model
 - i) Proactive (8)
 - ii) Governed (8)

10.
 - i) Design a Typical Data Governance Organization Structure (8)
 - ii) Design the 3 different approaches in Data Quality Management (8)

UNIT IV

PART – A

1. Define Organization systems.

A social unit of people that is structured and managed to meet a need or to pursue collective goals. All organizations have a management structure that determines relationships between the different activities and the members, and subdivides and assigns roles, responsibilities, and authority to carry out different tasks. Organizations are open systems--they affect and are affected by their environment.

2. List Organization schemes.

- Alphabetical schemes
- Chronological schemes
- Geographical schemes

3. Define Organizing Information.

Organizing information is at the heart of information science and is important in many other areas as well. In bibliographic and similar information systems it involves classification as well as the description of documents or other entities; in database management it is known as data modelling; in artificial intelligence, as knowledge representation for expert systems, natural language understanding, and other purposes; in psychology, as the structure of memory and cognition; in linguistics, as syntax and semantics and structure of discourse; in technical writing, as the structure of a composition; in biology it is used on two levels: in the classification of organisms and in the study of information transferred through genes.

4. List the exact Organization schemes.

Exact organization schemes objectively divide information into mutually exclusive sections. These systems comparatively are easy for information architects to create and categorize content within. :

- **Alphabetical schemes** make use of our 26-letter alphabet for organizing their contents. For this type of scheme to be successful, it is important that the content labeling matches the words that users are looking for. Sometimes, alphabetical schemes in the form of an A-Z index serve as secondary navigational components to supplement content's findability that is otherwise organized.
- **Chronological schemes** organize content by date. For these schemes to be successful there must be agreement about when the subject of the content took place.
- **Geographical schemes** organize content based on place. Unless there are border disputes, this type of scheme is fairly straightforward to design and use.
-

5. Name any organizational structures.

- Functional Structure
- Divisional Structure
- Matrix

6. Define heterogeneity.

Heterogeneity is a word that signifies diversity. A classroom consisting of people from lots of different backgrounds would be considered having the quality of heterogeneity. The prefix hetero- means "other or different," while the prefix homo- means "the same".

7. Summarize types of navigation systems.

- Automotive navigation system.
- Marine navigation system.
- Global Positioning System, a group of satellites and computers that can provide information on any person, vessel, or vehicle's location via a GPS receiver.

8. Give the meaning of Tuning and Tweaking.

Process in which one or more parameters of a device or model are adjusted upwards or downwards to achieve an improved or specified result.

9. Discuss Content models.

A content model is a representation of the types of content and their inter-relationships. For example, a car dealership may have content types for Vehicle, Dealer, and Manufacturer because these are first class citizens in the business domain. A restaurant have Recipes, Chef, Menu and Venue. The aim of the content model is to identify these content types, model their relationships and provide a shared semantic understanding of content from sponsor to developer.

10. Illustrate Hypertext.

A software system allowing extensive cross-referencing between related sections of text and associated graphic material.

11. Classify types of hierarchies.

1. Balanced hierarchy
2. Unbalanced hierarchy
3. Ragged hierarchy

12. Explain Social Navigation.

- The transformation of an interface (usually Web based) by using the actions of visitors.
- The process of guiding activities aimed at determining our position and planning and following a specific route based on what other people has done or what other people have recommended doing.

13. Deduce the features of hypertext structure.

- Materials are arranged in a logical, non-linear, order.
- Hypertext documents can include color pictures, animation, Sketchpad documents, forms and scripts, Java programs, and of course links to other relevant hypertext materials. However, students may not be able to use all (or any) of these features of the Web from home computers, or from printed out copies of the materials.
- Students are exposed to Web publishing and browsing in the act of reading the instructional materials, but may become sidetracked or lost if too many links to outside materials are available.

PART – B

1. Tell the following Principles of Information Architecture in detail

- i) The principle of objects (4)
- ii) The principle of choices (4)
- iii) The principle of disclosure (4)
- iv) The principle of exemplars (4)

2. Explain the following in detail

- i) Brainstorming with White Boards and Flip charts (8)
- ii) Metaphor Exploration (8)

3. Describe the granularity of content in detail. (16)

4. Describe the following in detail

- i) High level architecture Blueprints (8)
- ii) Web Based Prototypes (8)

5. Discuss the following in detail

- i) Navigation System Choices (6)
- ii) Index Terms (6)
- iii) Iconic Labels (4)

6. Discuss Principles of Information Architecture

- i) The principle of front doors (4)
- ii) The principle of multiple classification (4)
- iii) The principle of focused navigation (4)

- iv) The principle of growth (4)
- 7. Illustrate the following in detail about Embedded Navigation Systems
 - i) Global Navigation Systems (6)
 - ii) Local Navigation Systems (6)
 - iii) Contextual Navigation Systems (4)
- 8. Illustrate the organizational structures in detail (16)
- 9. Briefly explain the following in detail
 - i) Personalisation and Customization (8)
 - ii) Visualization and Social Navigation (8)
- 10. Explain the following in detail
 - i) Implementing Embedded Navigation (8)
 - ii) Supplemental Navigation System (8)
- 11. Explain the following in detail
 - i) Contextual Links (8)
 - ii) Headings (8)
- 12. Design the following for label design
 - i) Guidelines (8)
 - ii) Related issues (8)

UNIT V

PART – A

1. Define Data retention policy.

Data retention defines the policies of persistent data and records management for meeting legal and business data archival requirements; although sometimes interchangeable, not to be confused with the Data Protection Act 1998.

2. Define Sensitive Data.

Sensitive data encompasses a wide range of information and can include: your ethnic or racial origin; political opinion; religious or other similar beliefs; memberships; physical or mental health details; personal life; or criminal or civil offences.

3. Tell about Confidential/Regulated Data.

Confidential Data is a generalized term that typically represents data classified as Restricted, according to the data classification scheme defined in this Guideline. This term is often used interchangeably with sensitive data.

4. Describe the Data Retention (EC Directive) Regulations 2009.

Data Retention (EC Directive) Regulations 2009 came into force on 6 April, in the face of opposition from privacy campaigners and serious questions from lawyers. Billed as a vital tool in the fight against terrorism and other crime, the regulations stand accused of being yet another snoopers' charter.

5. Name any two exemptions for freedom of information act 2000.

- ABSOLUTE exemptions
- NON-ABSOLUTE exemptions

6. Define ILM.

Information life cycle management (ILM) is a comprehensive approach to managing the flow of an information system's data and associated metadata from creation and initial storage to the time when it becomes obsolete and is deleted.

7. Discuss key benefits of PLCM.

1. Reduced time to market
2. Reduced market entry costs
3. More efficient and profitable distribution channels
4. Higher return on investment from promotional campaigns
5. Extend the lifetime of your product by adapting your approach as it moves through the lifecycle
6. Orderly and profitable end of life product management

8. Give the Challenges in Big Data Testing.

- Huge Volume and Heterogeneity
- Understanding the Data
- Dealing with Sentiments and Emotions
- Lack of Technical Expertise and Coordination
- Stretched Deadlines & Costs

9. Discuss Sensitivity Analysis.

A sensitivity analysis is a technique used to determine how different values of an independent variable impact a particular dependent variable under a given set of assumptions.

10. Demonstrate Big data testing strategy.

Testing Big Data application is more a verification of its data processing rather than testing the individual features of the software product. When it comes to Big data testing, performance and functional testing are the key.

11. Illustrate Data migration testing.

Data Migration is one of the most challenging initiatives in IT Industry. These projects yield high business benefits (such as cost savings, increased productivity and improved data manageability) but at the same time they tend to involve a high level of risk due to the volume and criticality/complexity of the data.

12. Analyse tools used in big data scenarios.

- Jaspersoft BI Suite
- Pentaho Business Analytics
- Karmasphere Studio and Analyst
- Talend Open Studio
- Skytree Server
- Tableau Desktop and Server

13. Classify testing of Big Data.

- Architecture Testing
- Performance Testing

14. Explain parameters for performance testing.

1. No.of users the application can support
2. Resposne time what the application can give under the sustainable load

PART – B

1. Describe Data retention policies in detail.
2. Tell about the following Sensitive Data Handling Procedures
 - i) Security Classifications (6)
 - ii) Use, Transmission and Disposal of Records, Computers and Media (10)
3. i) Describe 3 tiers of data classification in detail (8)
 - ii) Describe the predefined types of restricted information assets(8)
4. Summarize the Information Security and the internet (16)
5. Describe the ten signs of Big Data and Analytics testing in detail (16)
6. i) Discuss about archiving data with Hadoop(8)
 - ii) Discuss about Cost Reduction toolbox (8)
7. Illustrate confidential data in detail with examples (16)
8. Explain Challenges in Big Data Testing (16)
9. Prepare a case study for handling confidential information (16)